Predicting Mental Health Care Outcomes Using Linear Regression

Anusha Yarasu

Northwood University

Abstract

The biggest part of public health is mind wellness and looking ahead at mind health results can really help in making good programs. This study looks at using straight line math to predict mind health effects based on actions and background info. Many things such as age, money, and ways of living are part of the data set. A key piece of public health is mind wellness; thinking about mind outcomes can help a lot in creating good plans. This study investigates how straight-line math works for predicting those mental effects too. Many things like age, money, and life choices are in the data set. The outcomes of preprocessing the statistics and using a linear regression model indicate that, with a mean squared error of 0.390, the version accounts for approximately 37.9% of the variation in intellectual fitness rankings. Although there is capacity for improvement inside the version's overall performance, it offers insightful facts approximately the variables influencing mental health results, that can guide mental health interventions and policy. As revealed by the data preprocessing results and the application of a linear regression model, the model captures about 37.9% of the fluctuation in the mental fitness scores, with a mean squared error of 0.390%. Although the model could be made more efficient, it conveys valuable information about the aspects that impact the mental fitness of people, which in turn can be used as a guide for the development of mental fitness programs and policy.

Introduction

Mental health problems are recognized by the World Health Organization (WHO) as the major cause of the loss of industrial productivity and the leading European Union's (EU) socio-economic challenge. Most healthcare professionals believe that the problem is getting worse and should be dealt with swiftly. Preventions usually decrease the incidents of chronic diseases. Through growing concern over the social and economic costs of mental diseases, mental health is a crucial component of total health. Policy interventions and the allocation of resources to high-risk persons can both benefit from the prediction of mental health outcomes. The aim of this research is to commit linear regression to build a predictive model for mental health. These are the demographic and behavioral demographic factors that are essential for the student of mental health-related elements to have grasp and get a good outcome of the models that are the basis of the model's predictions. The main purpose of this research is to provide strategies that can influence national health policymaking by pointing out the dominant factors.

Objective and Goal:

The main objective of our research is to be able to predict mental health scores from behavioral and demographic data. The most important job of the research is to identify the effects of the most important mental health variables using a Linear Regression model. That information, as of the present, is still not available. The pros of the new model will be its focus on key mental health factors, possible correctives and its use for treatment and decision-making.

Methodology

Dataset:

The dataset of the project combines behavioral and demographic characteristics such as: Demographics: income level, age, gender, etc. Behavioral: Regularity of exercise, degree of social engagement, usage of drugs, etc. To address missing values and ultimately standardize the variables for the linear regression model, the dataset was processed.

Preprocessing:

Imputation of missing values, transformation of continuous variable values for normality, and categorization of coding for qualitative variables were all practices included in the pre-processing phase. To make the model be error-resistant, which might cause the assumption of linear to be wrong, we looked for and treated any outliers properly.

Model Development:

Despite its strong performance in revealing continuous outcomes, linear regression is still selected as the predictive model. The following equation was utilized:

$$Y = \beta 0 + \beta 1 X 1 + \beta 2 X 2 + ... + \beta n X n + \epsilon$$

For example, $Y=\beta 0 +\beta 1 X 1 +\beta 2 X 2 +...+\beta n X n +\epsilon$

The mental health score is represented by Y, and the predictor variables are Xn. calculation of the model, one of the things that this social media company tries to do is to figure out the link between a person's mental health (i.e., he/she can behave healthily) and his/her online behavior.

Model Evaluation:

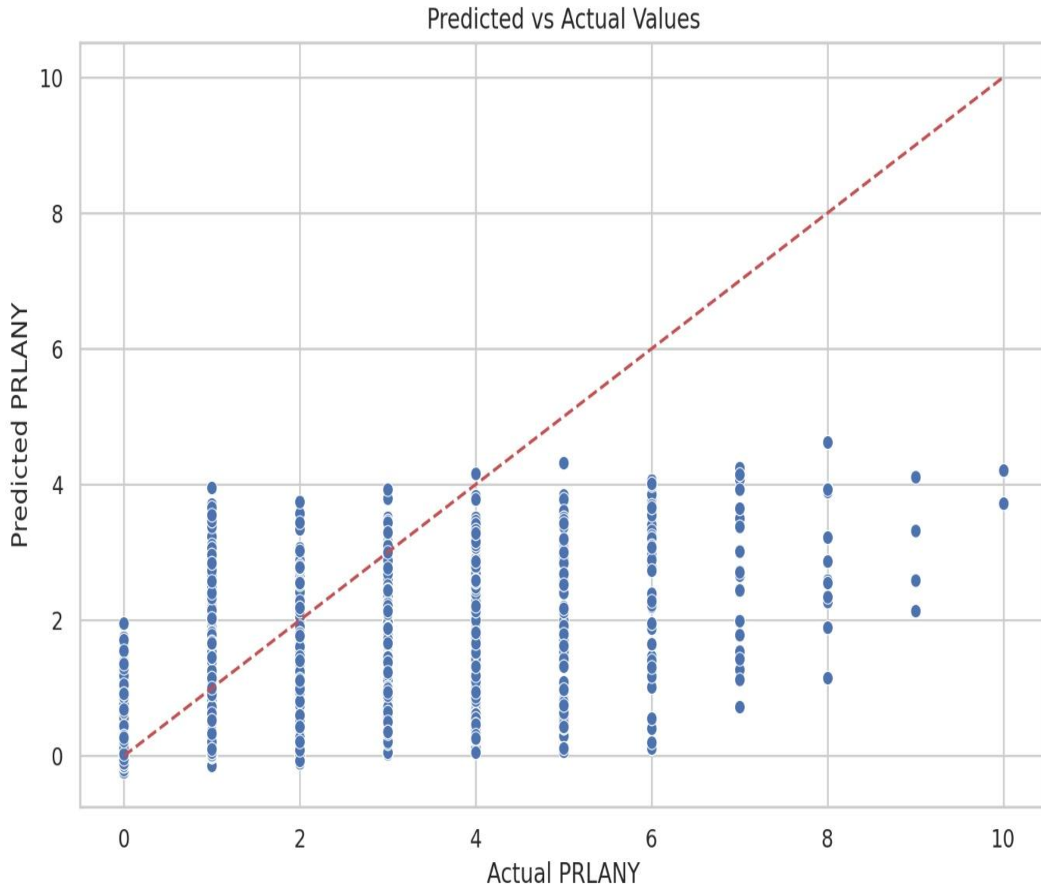
The measurement data is shown to illustrate how the model is validated: R squared = 0.379 indicates that 37.9% of the variability in people's mental health ratings is explained by the model. The value is 0.390. Aside from the procedure of cross-validation that was done to support the model's generalizability.

Results

The linear regression model had been 37.9% accurate, and it's forecasting outcomes for mental health, with an R-squared of 0.379.The mean squared error, with an average of 0.390, is the numerical difference between the expected and actual mental health ratings.

Visualizations:

The visualization involved key graphs that showed the performance of the model: A plot of the residuals: This graph not only provided a visual analysis of the homoscedasticity assumption but also showed the errors' distribution. Plot prediction versus actual: To demonstrate the model's appropriateness, actual and estimated mental health scores were contrasted.



The scatter plot illustrates the relationship between predicted and actual values of the variable PRLANY. The points on the graph represent individual data points, with the x-axis showing the actual values and the y-axis showing the corresponding predicted values. An ideal model would have all points falling directly on a diagonal line, indicating perfect agreement between predictions and actual values. In this case, while there is a general trend towards agreement, there is some scatter around the line, suggesting that the model's predictions are not entirely accurate. However, the overall pattern suggests a positive correlation between predicted and actual values, indicating that the model is able to capture the underlying trend in the data.

Discussion

Through the examination of the performance of the linear regression model, it is shown that, despite its efficiency in finding the meaningful associations between personality traits and demographic characteristics and health, its predictability is relatively moderate. The R2 value at 0.379 reveals that the inclusion of extra variables or the specification of nonlinear relations on the model could maximize the predictive capability.

Challenges:

Data quality assurance was of utmost concern for the project. One way to avoid bias in data collections is to handle the missing variables carefully. Similarly, if the linearity assumption is violated for some variables, model performance will be affected.

Ethical Considerations:

Predictive models used in mental health are required to address ethical concerns such as the privacy of data and the likelihood of predictions based on demographic characteristics being biased. Both these factors have been under the spotlight since the use and development of the model.

Conclusion

This study reveals that a linear regression model can be a foundation to predict mental health outcomes according to the behavior data and the demographic data. The fact that the model can explain only 37.9% of the variance in this study is its only shortcoming, however, it is a very good tool which we can use to know what influences people's mental health and which we can use for future studies or policies.

Future Work:

To effectively express the complexity of mental health data, future research could focus on data modeling methods like random forests or neural networks. More detailed data may also be added to improve the predictive accuracy.

<u>References</u>

Pearson, R., Pisner, D., Meyer, B., Shumake, J., & Beevers, C. G. (2018). A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychological Medicine*, *49*(14), 2330-2341.

Kaur, I., Kamini, K., Kaur, J., Gagandeep, S. P., & Gupta, U. (2024). Enhancing explainability in predicting mental health disorders using human–machine interaction. *Multimedia Tools and Applications*.